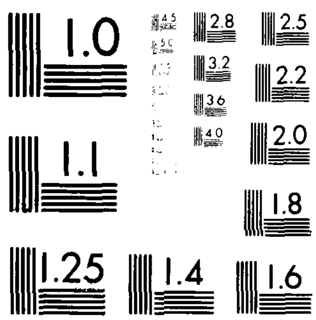
AD-A101 097 NAVAL UNDERWATER SYSTEMS CENTER NEW LONDON CT NEW LO--ETC F/G 12/1
IMPORTANCE SAMPLING FOR ESTIMATION OF SMALL PROBABILITIES.(U)
MAR 81 A H NUTTALL
UNCLASSIFIED NUSC-TR-6449 NL

1 of 1

END
DATE
FILMED
7-81
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

NUSC Technical Report 6449
23 March 1981

Importance Sampling for Estimation of Small Probabilities

Albert H. Nuttall
Surface Ship Sonar Department

Naval Underwater Systems Center
Newport, Rhode Island / New London, Connecticut

Preface

This research was conducted under NUSC IRIED Project No. A75205, Subproject No. ZR0000101, Applications of Statistical Communication Theory to Acoustic Signal Processing, Principal Investigator Dr. Albert H. Nuttall (Code 3302), Program Manager CAPT. David F. Parrish, Naval Material Command (MAT 08L).

The Technical Reviewer for this report was Norman Owsley (Code 3211).

Reviewed and Approved: 23 March 1981

W. A. Von Winkle

W. Von Winkle
Associate Technical Director for Technology

The author of this report is located at the
Naval Underwater Systems Center, New London
Laboratory, New London, Connecticut 06320.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (14) NUSC-~~TR~~-6449	2. GOVT ACCESSION NO. AD-A101 097	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Importance Sampling for Estimation of Small Probabilities		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Albert H. Nuttall	(9) Technical rept.	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Material Command (MAT 08L) Washington, D.C. 20360		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS A75205
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Underwater Systems Center New London Laboratory New London, CT 06320	(11)	12. REPORT DATE 23 March 1981
		13. NUMBER OF PAGES 41 (12) 47
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Distribution of Estimate	Simulation	Variance Reducing
False Alarm Probability	Unbiased Estimate	
Importance Sampling	Variance of Estimate	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The use of importance sampling for estimation of small probabilities is illustrated by means of a signal-detection example. In particular, false alarm probabilities in the 10^{-7} range are accurately estimated by means of only 1000 independent trials. The fundamental variance-reducing capability of importance sampling is explored and used to significantly improve the performance of the probability estimate for the detection example considered. Guidelines for choosing good data-generation algorithms are presented.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

405918 xll

Table of Contents

A

List Of Illustrations

List of Tables

List of Symbols

N	number of samples is $N + 1$
x_n	n-th data value
X	observation vector
$p_o(X)$	probability density function for noise-only
β	unknown noise power level measure
$p_1(X)$	probability density function for signal-present
γ	unknown signal-plus-noise power level measure
V	threshold value
P_{FA}	probability of false alarm
M	number of components of X
$z, g(X)$	processor output
$p(X)$	probability density function of X
P	probability of $z = g(X)$ exceeding threshold V
R_V	region of X space where $z = g(X) > V$
$X^{(i)}$	observation vector on i-th trial
T	total number of trials
$U(\ )$	unit step function
h_1, h_2, h_3, h_4	counting functions
$\alpha_1, \alpha_2, \alpha_3, \alpha_4$	estimates of probability P
$E\{\ \}$	ensemble average value
$SD\{\ \}$	standard deviation
$V\{\ \}$	variance
$p^*(X)$	alternative probability density function of X
K	scaling factor of potential-signal sample
$e_M(x)$	partial exponential power series
R	region of X space where $p(X) > 0$

IMPORTANCE SAMPLING FOR ESTIMATION OF SMALL PROBABILITIES

INTRODUCTION

One method of describing the capability of a signal processing system is through its false alarm and detection probabilities for detection applications, or in terms of its error probabilities for communication applications. When these probabilities are not analytically available, simulation can often be employed to estimate them. However, for very small false alarm or error probabilities, it may not be possible, via direct simulation, to conduct enough independent trials to realize reliable estimates with sufficient stability.

This apparent shortcoming is not an inherent limitation of estimation, but is due instead to the discrete counting procedure often adopted in direct simulation. It is possible to remedy this situation by using a "continuous" counting procedure, whereby the result of each individual trial can take on a continuum of values, the range of which can include arbitrarily small probabilities. In addition, the variance of the resultant estimate can be reduced to arbitrarily-small values, even for a limited number of independent trials, *provided* that the proper data-generation method is used.

This technique, known as importance sampling (reference 1), will be explained and explored here by means of a particular signal-processing example presented by Hansen (reference 2). In addition, the fundamental variance-reducing capability will be investigated and used to derive a better data-generation technique. Guidelines for choosing good data-generation algorithms will also be presented.

SIGNAL DETECTION EXAMPLE

The importance sampling technique will be explained by means of the following signal detection example. Suppose that we observe N+1 samples $\{x_n\}$ of some random process. Let the probability density function (PDF) of the observation vector

$$X = (x_1, x_2, \ldots , x_{N+1}) \tag{1}$$

for noise-only be denoted by

$$p_o(X) = \prod_{n=1}^{N+1} \left\{\frac{1}{\beta} \exp\left(-\frac{x_n}{\beta}\right)\right\} \text{ for all } x_n > 0 \quad , \tag{2}$$

where β is unknown; that is, the power level of all samples is identical but is unknown. Also, let the PDF of X for signal present be

$$p_1(X) = \prod_{n=1}^{N} \left\{ \frac{1}{\beta} \exp\left(-\frac{x_n}{\beta}\right) \right\} \frac{1}{\gamma} \exp\left(-\frac{x_{N+1}}{\gamma}\right) \text{ for all } x_n > 0 \ , \tag{3}$$

where γ is unknown, but $\gamma > \beta$; that is, the power level of the potential-signal sample x_{N+1} is larger, but is also unknown.

The generalized likelihood ratio is derived in appendix A and leads to the threshold comparison test

$$\frac{x_{N+1}}{\frac{1}{N}\left(x_1 + x_2 + \ldots + x_N\right)} \underset{H_0}{\overset{H_1}{\gtrless}} V \ . \tag{4}$$

The false alarm probability is given by the probability that the left side of (4) exceeds V when p_0 in (2) is the prevalent PDF of X. This is the example considered in reference 2, equations (4)-(7).

Analytic evaluation of the false alarm probability for test (4) and PDF (2) is readily accomplished in equations (A-9)-(A-11) of appendix A:

$$P_{FA} = \frac{1}{(1 + V/N)^N} \ . \tag{5}$$

The exact value of β in (2) is irrelevant in test (4), since the left side of (4) is independent of absolute levels; hence P_{FA} depends only on the number N of noise-only samples and the threshold V. This is called a constant false alarm receiver, since the absolute noise level need not be known in order to realize a specified false alarm probability. In fact, (5) can be solved directly for the threshold required as

$$V = N\left(P_{FA}^{-1/N} - 1\right) \ , \tag{6}$$

in terms of the specified or desired P_{FA} and the number of samples N. Since the value of β is irrelevant in test (4), we will set $\beta = 1$, henceforth, without loss of generality.

DEFINITION OF PROBLEM

The general situation of interest is depicted in figure 1. X is an observation vector of M components, with known PDF p(X). The processor takes this collection of M samples, X, and emits a single quantity, z, according to transformation

$$z = g(X) \ , \tag{7}$$

which is compared with threshold V. The known quantities here are the input PDF p(X), the (nonlinear) transformation g(X), and the threshold V. There may be statistical dependence between the components $\{x_n\}$ of the observation. Also, the input PDF and the transformation are arbitrary but fixed. (In the example of the previous section, g(X) is given by the left side of (4), and p(X) is given by (2).)

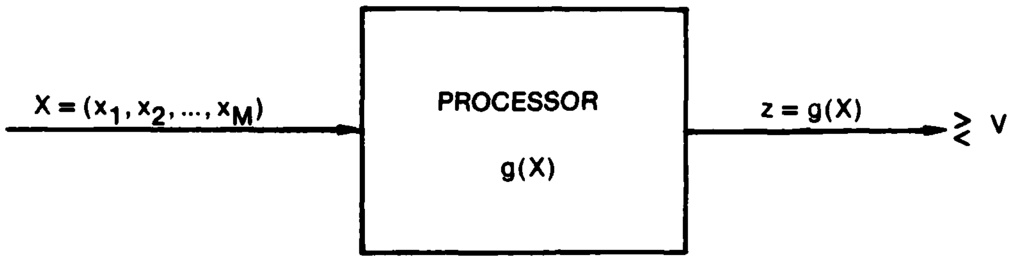


Figure 1. General Processor of Observation X

We want to evaluate the threshold-crossing probability (exceedance probability)

$$P \equiv \text{Prob}\left\{ z > V \right\} = \text{Prob}\left\{ g(X) > V \right\} = \int_{R_v} dX \; p(X) \quad , \tag{8}$$

where R_v is defined as the region of X space where $g(X) > V$. If $p(X)$ is the PDF $p_0(X)$ for noise-alone at the processor input, then P is the false alarm probability, whereas if $p(X)$ is the PDF $p_1(X)$ for signal-present, then P is the detection probability. We shall be concerned with the former case where the false alarm probability is very small.

There are at least two major analytical difficulties with the problem statement in (8): (a) explicit determination of the region R_v may be very difficult to achieve, especially for large M; (b) evaluation of P via the integral in (8) may be very difficult to carry out, even if R_v is explicitly specified. For large M, these analytical difficulties are virtually always insurmountable, except for special regions R_v and special PDFs. Accordingly, it is frequently necessary to resort to a simulation to estimate P. In this report, we will consider the performance of: a direct simulation; a modified simulation indicated by importance sampling; and some additional simulations indicated by the optimum PDF for importance sampling.

DIRECT SIMULATION

Since the PDF of observation X is known, we presume that we can generate data subject to these statistics. In particular, suppose we generate, according to PDF p, the i-th observation vector $X^{(i)}$, statistically independent of $X^{(j)}$ for $j \neq i$, for a total of T trials; i.e., $1 \leq i \leq T$. Now define the unit step function

$$U(y) = \begin{Bmatrix} 1 \text{ for } y > 0 \\ 0 \text{ for } y < 0 \end{Bmatrix} . \tag{9}$$

Then we define our counting function on the i-th trial as

$$h_1\left(X^{(i)}\right) = U\left(g\left(X^{(i)}\right) - V\right) = \begin{Bmatrix} 1 \text{ for } X^{(i)} \in R_V \\ 0 \text{ for } X^{(i)} \notin R_V \end{Bmatrix} . \tag{10}$$

That is, the result of the i-th trial is 1 or 0, depending on whether the threshold V is exceeded or not, respectively. Finally, the estimate of the desired probability P is furnished by the average of the counting function over the T independent trials:

$$\alpha_1 \equiv \frac{1}{T} \sum_{i=1}^{T} h_1\left(X^{(i)}\right) . \tag{11}$$

Observe that we use the known quantities p(X), g(X), and \ each trial (10).

This estimate is unbiased, because

$$E\{\alpha_1\} = E\{h_1(X)\} = \int dX\ p(X)\ h_1(X) = \int_{R_V} dX\ p(X) = P . \tag{12}$$

Here we used the facts that each observation $X^{(i)}$ was generated according to PDF p, that h_1 is given by (10), and relation (8).

The PDFs of random variables h_1 and α_1 are depicted in figure 2. The values for the areas of the impulses in the PDF for α_1 are given by the binomial quantity

$$Q_k = \binom{T}{k} (1 - P)^{T-k}\ P^k \text{ at } \alpha_1 = \frac{k}{T} , \text{ for } 0 \leq k \leq T , \tag{13}$$

since all T trials are independent. The mean value of each of the random variables is also indicated in the figure, and serves to point out the fundamental limitation of such a direct simulation. Specifically, the result h_1 of a trial can never equal the desired quantity P, but can only take on the values 0 and 1. The averaging of T trials helps considerably, but if P is significantly less than 1/T, the estimate yielded by random variable α_1 is inadequate since it is either too small (0) or too large (1/T, 2/T,...).

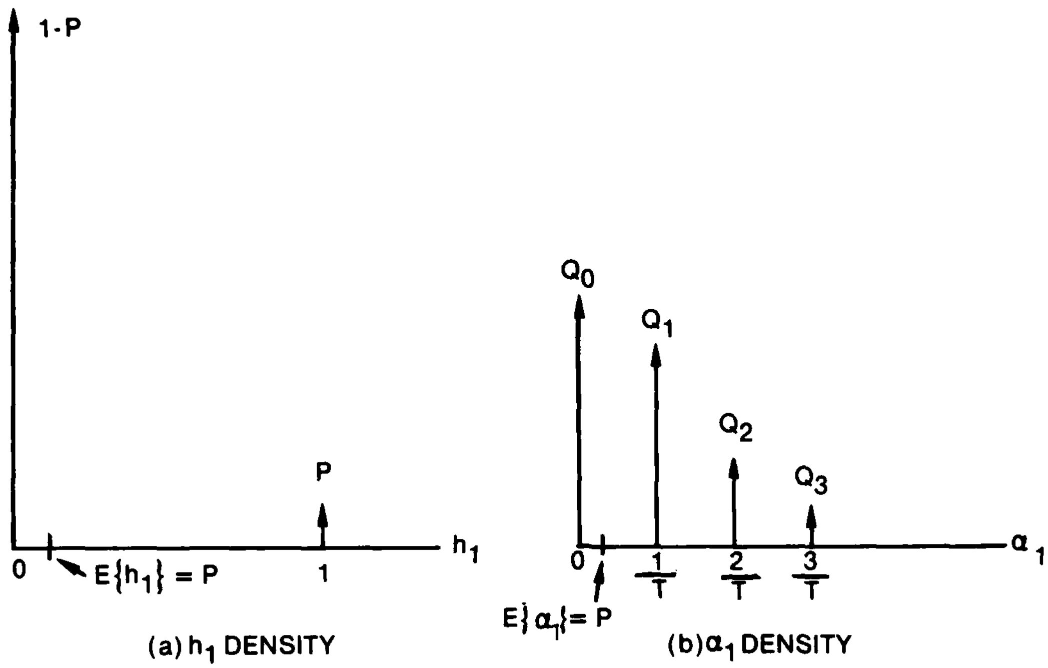


Figure 2. Probability Density Functions for h_1 and α_1

The result of a simulation by means of counting function h_1 in (10), for the signal detection example in (4),

$$g(X) = \frac{x_{N+1}}{\frac{1}{N}(x_1 + x_2 + \ldots + x_N)} \gtrless V \quad , \tag{14}$$

with $N = 32$ and $T = 1000$, is presented in figure 3. The exact result in figure 3 is that already given by (5) and appendix A. The simulation via h_1 was conducted only at the integer values of V, and is observed to limit at $1/N = 10^{-3}$ before jumping to 0. None of the values of P for $V > 8$ can be accurately estimated via this direct simulation.

The variance of h_1 is $P(1-P)$, and that of α_1 is $P(1-P)/T$, since the T trials are independent. The ratio of the standard deviation of α_1 to its mean is $((1-P)/(PT))^{1/2}$, which is small only if T is significantly larger than $1/P$. As a comparison case against which future estimates will be compared, we find that for

$$N = 32, \; V = 8, \; \beta = 1, \; T = 1000, \tag{15}$$

we have statistics

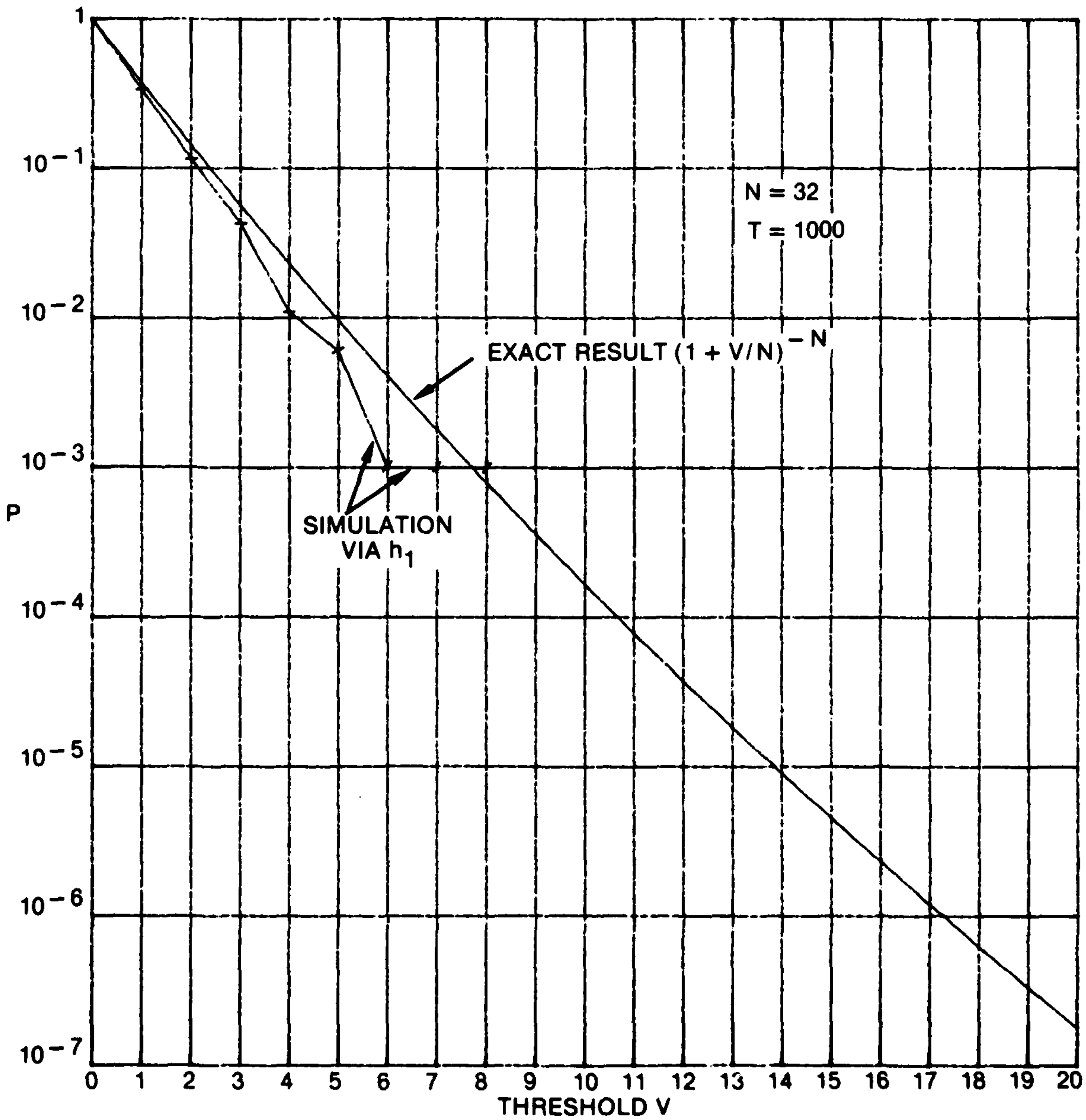


Figure 3. Direct Simulation Result

$$E\{h_1\} = 0.000792, \ SD\{h_1\} = 0.0281,$$

$$E\{\alpha_1\} = 0.000792, \ SD\{\alpha_1\} = 0.000890,$$

$$P = 0.000792, \ Q_o = 0.453, \ Q_1 = 0.359, \ Q_2 = 0.142, \ Q_3 = 0.038, \ldots \tag{16}$$

Here, SD denotes the standard deviation. Thus, the standard deviation of estimate α_1 is still greater than its mean value, even though an average of 1000 trials has been employed. The reason for this behavior is because h_1 is such a poor indicator of its mean value; in fact, its standard deviation is 35.5 times greater than its mean value. An alternative counting function to h_1 that is more closely peaked around its average value must be found.

IMPORTANCE SAMPLING

Suppose we generate observation X according to alternative PDF p*(X), instead of the originally specified p(X). Also let us use counting function

$$h(X) \equiv \frac{p(X)}{p^*(X)} U(g(X) - V) \tag{17}$$

instead of (10). Observe that the same known quantities, p(X), g(X), and V are involved in (17), in addition to the yet-to-be-specified PDF p*(X). Also, h is no longer restricted to just the values 0 or 1, as (10) was, due to the scaling p/p*. The transformation of interest, g(X), and the threshold V are not changed in any way.

The estimate of P is obtained by performing T independent trials as earlier, and averaging the results:

$$\alpha \equiv \frac{1}{T} \sum_{i=1}^{T} h\left(X^{(i)}\right) , \tag{18}$$

where the i-th observation $X^{(i)}$ is generated according to alternative PDF p*(X), not p(X).

The random variables h and α are unbiased estimators of P, since

$$\begin{aligned} E\{h(X)\} &= \int dX \ p^*(X) \ h(X) \\ &= \int dX \ p(X) \ U(g(X) - V) = \int_{R_V} dX \ p(X) = P \ . \end{aligned} \tag{19}$$

Observe in the first line of (19) that the average of h must be performed according to PDF p*(X), not p(X), since the data X was generated according to p*(X); we then employed (17), (10), and (12). The general nature of the PDF of counting function h in (17) is displayed in figure 4. There could still be a non-zero probability of getting

h = 0, depending on the choice of p*(X) in (17); however, this probability can be made much less than for a direct simulation. Also there is a distributed portion of the PDF, hopefully peaked near E{h} = P.

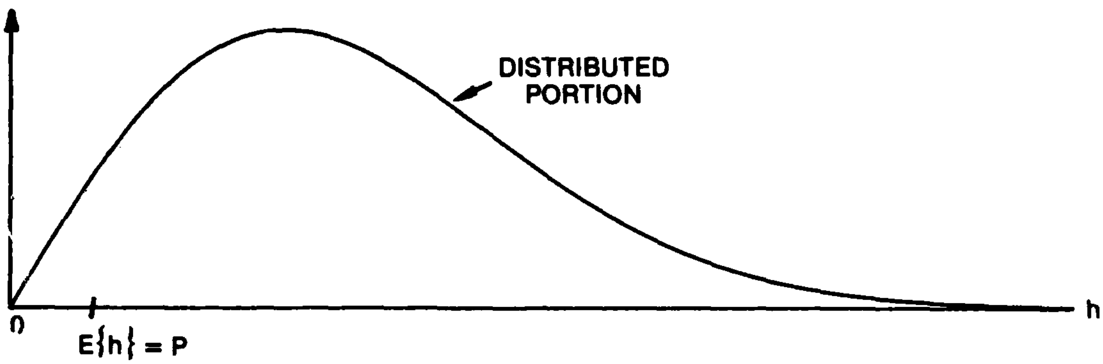


Figure 4. Probability Density Function for h

Since the T trials leading to estimate α in (18), of the probability P, are statistically independent, the variance of α is given by

$$V\{\alpha\} = \frac{1}{N} V\{h\} = \frac{1}{N}\left[E\{h^2\} - E^2\{h\}\right] \quad . \tag{20}$$

We have already evaluated E {h} in (19). The remaining average required in (20) is

$$E\{h^2\} = \int dX\ p^*(X)\ h^2(X) = \int dX\ \frac{p^2(X)}{p^*(X)}\ U(g(X) - V) \quad , \tag{21}$$

which depends on p* as well as p, g, and V; we have again averaged h^2 according to p* in (21), and used (17). Selection of p* for a minimum of (21) will be considered later.

SCALING OF POTENTIAL-SIGNAL SAMPLE

The first example of importance sampling that we consider is the one in reference 2, pp. 548-550. The alternative PDF, p*, is chosen so that inputs X, for which a large output z results in figure 1, are generated with an increased probability (reference 2, p. 546). Specifically, instead of the original PDF (with β = 1)

$$p(X) = \prod_{n=1}^{N+1} \{p(x_n)\} \text{ with } p(x) = e^{-x} \text{ for } x > 0 \quad , \tag{22}$$

we use, for data generation, the alternative PDF

$$p^*(X) = \prod_{n=1}^{N} \{p(x_n)\} \frac{1}{K} p\left(\frac{x_{N+1}}{K}\right) \quad \text{with } K > 1 \quad . \tag{23}$$

Thus the potential-signal sample, x_{N+1}, has been scaled by K and will more often lead to satisfaction of the threshold crossing in (4). Use of (22), (23), and (14) in (17) leads to counting function

$$h_2(X) \equiv K \exp\left(-x_{N+1}\left(1 - \frac{1}{K}\right)\right) U\left(x_{N+1} - \frac{V}{N} s\right) , \tag{24}$$

where

$$s \equiv \sum_{n=1}^{N} x_n \quad . \tag{25}$$

(If K = 1, (24) reduces to (10), the direct simulation case.) The corresponding estimate of P is given according to (18) as

$$\alpha_2 = \frac{1}{T} \sum_{i=1}^{T} h_2\left(X^{(i)}\right) \quad . \tag{26}$$

The result of a simulation via h_2 and α_2 in (24) and (26) is given in figure 5 for the comparison case cited in (15), with scaling factor K = 6. The contrast between figures 3 and 5 is very pronounced. Now estimates of P all the way down to 10^{-7} are possible via use of h_2, whereas previously, the direct simulation could not yield estimates less than $1/T = 10^{-3}$. Also, the standard deviation of the estimates in figure 5 is observed to be very small for the smaller values of V, although it gets larger as V increases. The program for figure 5 is given in appendix B; when K is set equal to 1, the results given in figure 3 occurred.

In order to determine the performance of this importance sampling procedure, and to ascertain if there is an optimum value of scaling K, we evaluate the variances of h_2 and α_2. In appendix C, the ν-th moment of h_2 is evaluated. In particular, there follows from (C-5),

$$E\left\{h_2^2\right\} = \frac{K}{2 - \frac{1}{K}} \frac{1}{\left[1 + \frac{V}{N}\left(2 - \frac{1}{K}\right)\right]^N} \quad . \tag{27}$$

Since

$$E\left\{h_2\right\} = P = \frac{1}{\left(1 + \frac{V}{N}\right)^N} \tag{28}$$

is independent of K (as expected), the variance of h_2 is minimized when (27) is minimized. There follows for the optimum value of scaling K, from (C-9),

$$K_o = \frac{1 + V + \frac{3V}{N} + \left(1 + \frac{2V}{N} + \left(V + \frac{V}{N}\right)^2\right)^{1/2}}{2\left(1 + \frac{2V}{N}\right)} \quad . \tag{29}$$

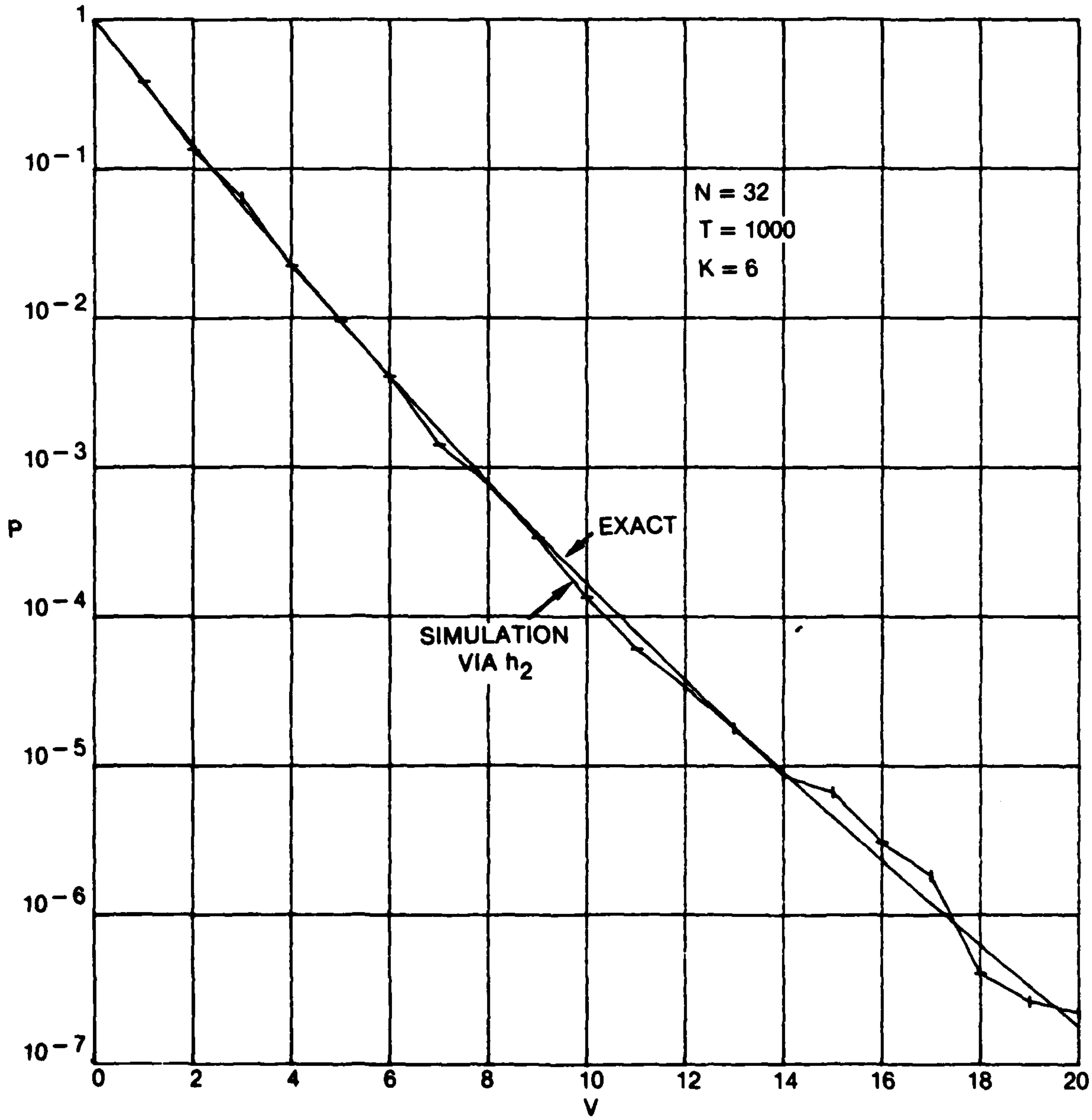


Figure 5. Simulation for Scaled Potential-Signal Sample

A table of the optimum scaling K_o is given below, along with the mean and minimum standard deviation of estimate α_2 defined in (26), for N = 32 and T = 1000. For V = 8, the minimum standard deviation of α_2 is 5.9 times smaller than its mean value, for example. This is far better than the situation in (16) for direct simulation, where the standard deviation of α_1 was greater than its mean. For larger V, i.e., low probability P, the minimum standard deviation is seen to become larger than the mean value P. Specifically this occurs for $V \geq 18$. Thus estimation of very low probabilities P via this particular importance sampling procedure is subject to significant error, even when scaling K is optimally selected. Of course, in practice, the optimum value of K will not be known, and a single value would likely be used for a range of values of V.

Table 1. Statistics of α_2 for N = 32, T = 1000

V	K_o	$P = E\{\alpha_2\}$	Min $SD\{\alpha_2\}$
2	2.45	1.44E-1	7.30E-3
4	3.86	2.31E-2	1.88E-3
6	5.04	4.09E-3	4.86E-4
8	6.03	7.92E-4	1.34E-4
10	6.87	1.66E-4	4.02E-5
12	7.59	3.75E-5	1.30E-5
14	8.22	9.05E-6	4.48E-6
16	8.77	2.32E-6	1.65E-6
18	9.25	6.28E-7	6.43E-7
20	9.68	1.79E-7	2.64E-7

Other important measures of the quality of counting function h_2 are furnished by its PDF and exceedance probability. These quantities are derived in appendix D. We find

$$\text{Prob}\left\{h_2 > H\right\} = \left(1 + \frac{V}{KN}\right)^{-N} \left[1 - e^{-A_1} e_{N-1}(A_1)\right]$$

$$- \exp\left(-\frac{a}{K}\right)\left[1 - e^{-A_2} e_{N-1}(A_2)\right] \quad , \tag{30}$$

where

$$a = \ln\left(\frac{K}{H}\right)\frac{K}{K-1} \; , \quad A_1 = a\left(\frac{N}{V} + \frac{1}{K}\right) \; , \quad A_2 = a\frac{N}{V} \quad , \tag{31}$$

and the partial exponential series is (reference 3, eq. 6.5.11)

$$e_M(x) \equiv \sum_{m=0}^{M} \frac{1}{m!} x^m \quad . \tag{32}$$

A limiting procedure on (30) shows that

$$\text{Prob}\left\{h_2 > 0\right\} = \left(1 + \frac{V}{KN}\right)^{-N} \tag{33}$$

and therefore that

$$\text{Prob}\left\{h_2 = 0\right\} = 1 - \left(1 + \frac{V}{KN}\right)^{-N} \quad . \tag{34}$$

This is the probability that counting function h_2 gives a zero output for observation X, as noted in the PDF in figure 4.

The PDF of h_2 is given in (D-11):

$$p(H) = \frac{\exp(-a/K)}{(K - 1)H}\left[1 - \exp\left(-\frac{N}{V}a\right) e_{N-1}\left(\frac{N}{V}a\right)\right] \quad \text{for } 0 < H \leq K \quad , \tag{35}$$

where a is still given by (31). A plot of this PDF is presented in figure 6 for N = 32, V = 8, and K = 6. Observe that the ordinate is a logarithmic scale. The area of the impulse at H = 0 is available from (34) as .729; this is far less than the impulse at $h_1 = 0$ in figure 2(a) with area 1-P = .999208 (see (16)). However, .729 is still a substantial probability to be associated with outputting a zero from the counting function h_2. The PDF in figure 6 is very skewed; in addition to the large impulse at H = 0, there is an integrable singularity at H = 0+. Although figure 5 indicates significant improvement over figure 3, the very skewed PDF in figure 6 indicates that a great deal more improvement should be possible through proper choice of alternative PDF p*.

Although we could calculate the PDF of α_1 explicitly (see figure 2 and eq. 13), this is not the case for α_2 here, as given by (26). We can easily calculate the cumulants of α_2, by means of (C-4), but calculation of the PDF would require the following numerical procedure: (a) take the Fourier transform of PDF (35), thereby obtaining the characteristic function of h_2; (b) raise this complex function to the N-th power; (c) take the inverse Fourier transform, thereby obtaining the PDF of α_2. Some relevant observations on this procedure are as follows: the cusp of (35) at H = 0+ should be subtracted out and transformed analytically; the Fourier transforms should be accomplished by employing FFTs; the cumulative distribution of α_2 could be found directly instead of its PDF (see references 4 and 5). We have not pursued this particular PDF, but rather have tried to improve on the counting function h_2 instead.

OPTIMUM DATA GENERATION

The fundamental idea behind importance sampling was presented earlier in (17)-(21). It was pointed out that minimization of the variance of the estimate α in (18) requires minimization of (21) by choice of the alternative PDF p*. This problem is undertaken in appendix E, with the result that the optimum PDF to use for data generation is

$$p_o(X) = \left\{ \begin{array}{ll} p(X)/P & \text{for } X \epsilon (R \cap R_V) \\ \\ 0 & \text{otherwise} \end{array} \right\} \quad , \tag{36}$$

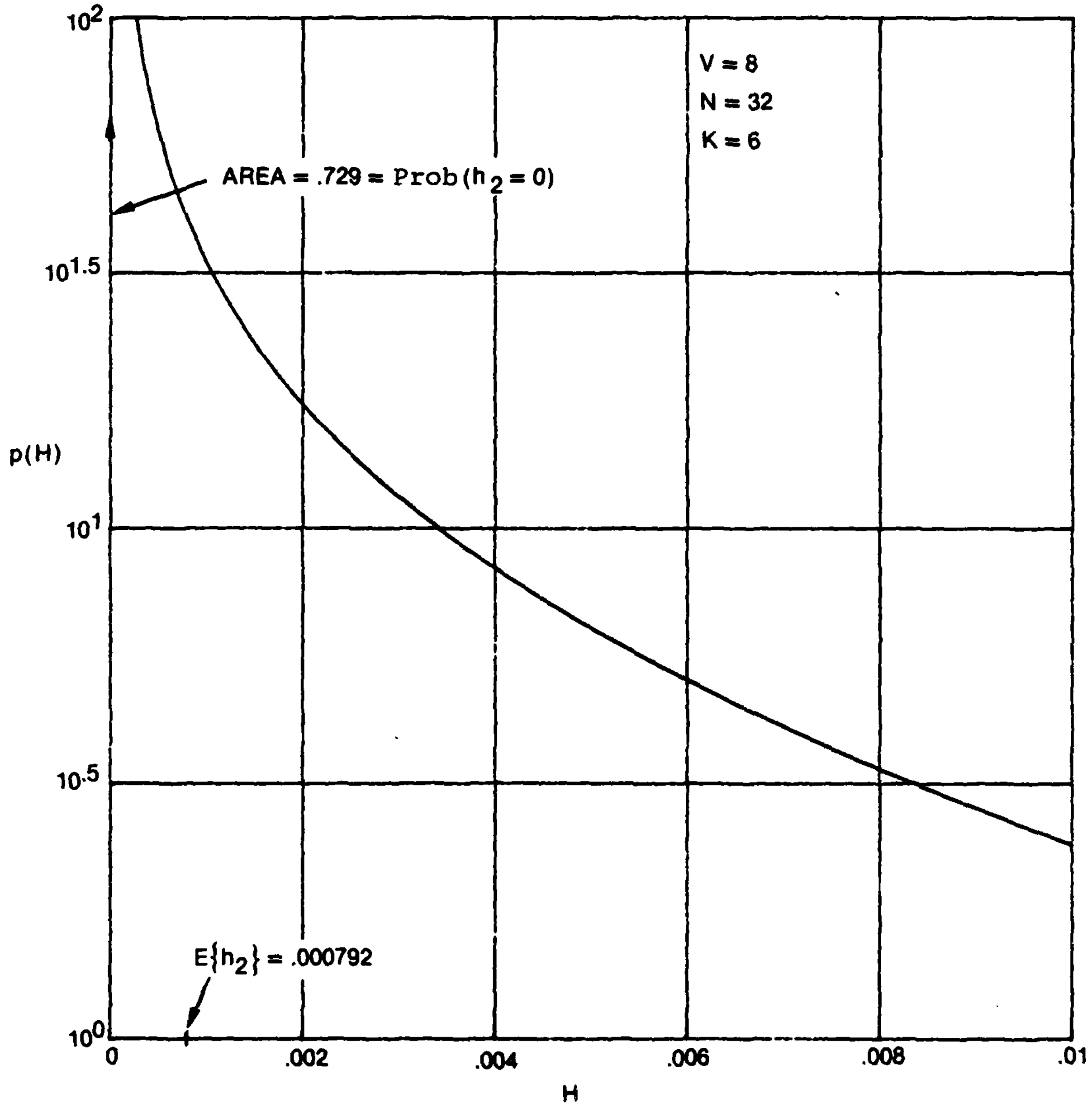


Figure 6. Probability Density Function for h_2

where the regions in X-space are described as

$$R : p(X) > 0;$$

$$R_V : g(X) > V \quad . \tag{37}$$

The form (36) for the optimum PDF is very illuminating. It says: generate X values only for which g(X) > V, and do it with a frequency proportional to the given PDF p(X). Furthermore, it says not to generate data X which leads to zero values for h, and not to generate data X which would not have been generated by the original PDF, p(X). Unfortunately, the value of the proportionality constant in (36) is P, the very quantity we are trying to estimate. In addition, determination of the region $R \cap R_V$ could be a very difficult analytical task.

The optimum counting function is shown in appendix E to be given by

$$h_o(X) = \begin{Bmatrix} P & \text{for} \quad X\varepsilon(R \cap R_V) \\ \\ 0 & \text{otherwise} \end{Bmatrix} \quad . \tag{38}$$

That is, every trial X generated according to (36) yields exactly the same value for the counting function; the value 0 in (38) is never encountered because $p_o(X)$ is zero for such data values X.

It follows that the variance of h_o (and the corresponding estimate α_o of P) is zero. Thus by proper choice of alternative PDF p*(X), we can reduce the variance of the estimation error to zero, for any fixed number of trials T. If instead of choosing p* exactly equal to p_o, we come reasonably close, then we shall realize the variance-reducing capability inherent in importance sampling (references 1, 2). Since the direct simulation approach always yields a zero output and is far from optimum, a significant improvement in estimation capability is often achieved with a minor change in the data-generating PDF; witness the results of the previous section which simply used a scaled version of the potential-signal sample and made no use of the optimum PDF for importance sampling. Even though direct usage of the optimum PDF in (36) is not feasible, it does furnish some good guidelines, as noted under (37). We shall use these guidelines in the next section to select some modified data-generation PDFs for the processor g(X) in (14) of interest here.

SOME ALTERNATIVE DATA GENERATION STRATEGIES

The original PDF p(X) is given in (22). Since the PDF and the test of interest, (14), involve $\{x_n\}_1^N$ only through their sum s defined in (25), we can rewrite this PDF as

$$p\left(s , x_{N+1}\right) = \frac{s^{N-1} \; e^{-s}}{(N - 1)!} \exp(-x_{N+1}) \quad \text{for} \quad s > 0 \; , \; x_{N+1} > 0 \quad , \tag{39}$$

and the test as

$$x_{N+1} \gtrless \frac{V}{N} s \quad . \tag{40}$$

A Shifted PDF

In keeping with the guidelines presented in the previous section, we take a shifted function for the conditional PDF:

$$p^*(s, x_{N+1}) = p^*(s) \cdot p^*(x_{N+1}|s)$$

$$= \frac{s^{N-1} e^{-s}}{(N-1)!} \cdot \exp\left[-\left(x_{N+1} - \frac{V}{N}s\right)\right] \quad \text{for } s > 0,\ x_{N+1} > \frac{V}{N}s \quad . \tag{41}$$

This PDF is non-zero only in $R_s \cap R$, as desired; however, it does not match the shape of (39) for all s, x_{N+1}, as (36) suggests. Then (17) yields counting function

$$h_3(X) = \exp\left(-\frac{V}{N}s\right) \quad \text{for} \quad x_{N+1} > \frac{V}{N}s > 0 \quad . \tag{42}$$

Furthermore, there is no need to generate x_{N+1} since it is not involved in h_3. Therefore we use (42) with the PDF for $p^*(s)$ as given in (41).

The exceedance probability of h_3 is immediately found from (42), (41), and (32):

$$\text{Prob}\ \{h_3 > H\} = \text{Prob}\left\{\exp\left(-\frac{V}{N}s\right) > H\right\} = \text{Prob}\{s < A_3\}$$

$$= \int_0^{A_3} ds\ \frac{s^{N-1} e^{-s}}{(N-1)!} = 1 - e^{-A_3}\ e_{N-1}(A_3) \quad \text{for } 0 \le H \le 1, \tag{43}$$

where

$$A_3 \equiv -\frac{N}{V} \ln H \quad . \tag{44}$$

The PDF of h_3 is available from (43) by taking a derivative with respect to H:

$$p(H) = \frac{N}{V(N-1)!} H^{\frac{N}{V}-1} \left(-\frac{N}{V} \ln H\right)^{N-1} \quad \text{for } 0 < H \le 1 \quad . \tag{45}$$

The range (0,1) for h_3 is immediately obvious from (42). We observe there is no impulse at H = 0 in the PDF (45) for h_3; in fact, (43) yields $\text{Prob}\{h_3 > 0\} = 1$. A plot of (45) is given in figure 7; although not peaked at $E\{h_3\} = P = .000792$, it is considerably better than figures 2 and 6 for h_1 and h_2, respectively.

The result of a simulation via counting function (42) for N = 32 and T = 1000 trials is given in figure 8. As done earlier, the simulation was conducted only at the integer values of V, and straight lines were drawn between these estimates. However, if the *same* random numbers constitute the set of observations $\{X^{(t)}\}_1^T$ for all the different threshold values V, as done in figure 8(a), a very misleading result and conclusion is possible; namely, it appears that there is a very small systematic error in the estimate α_3 of P. However, when *different* random numbers are used for the simulation at each value of V, the result in figure 8(b) correctly indicates an alternating but growing estimation error at the lower probabilities. Since in practice, the

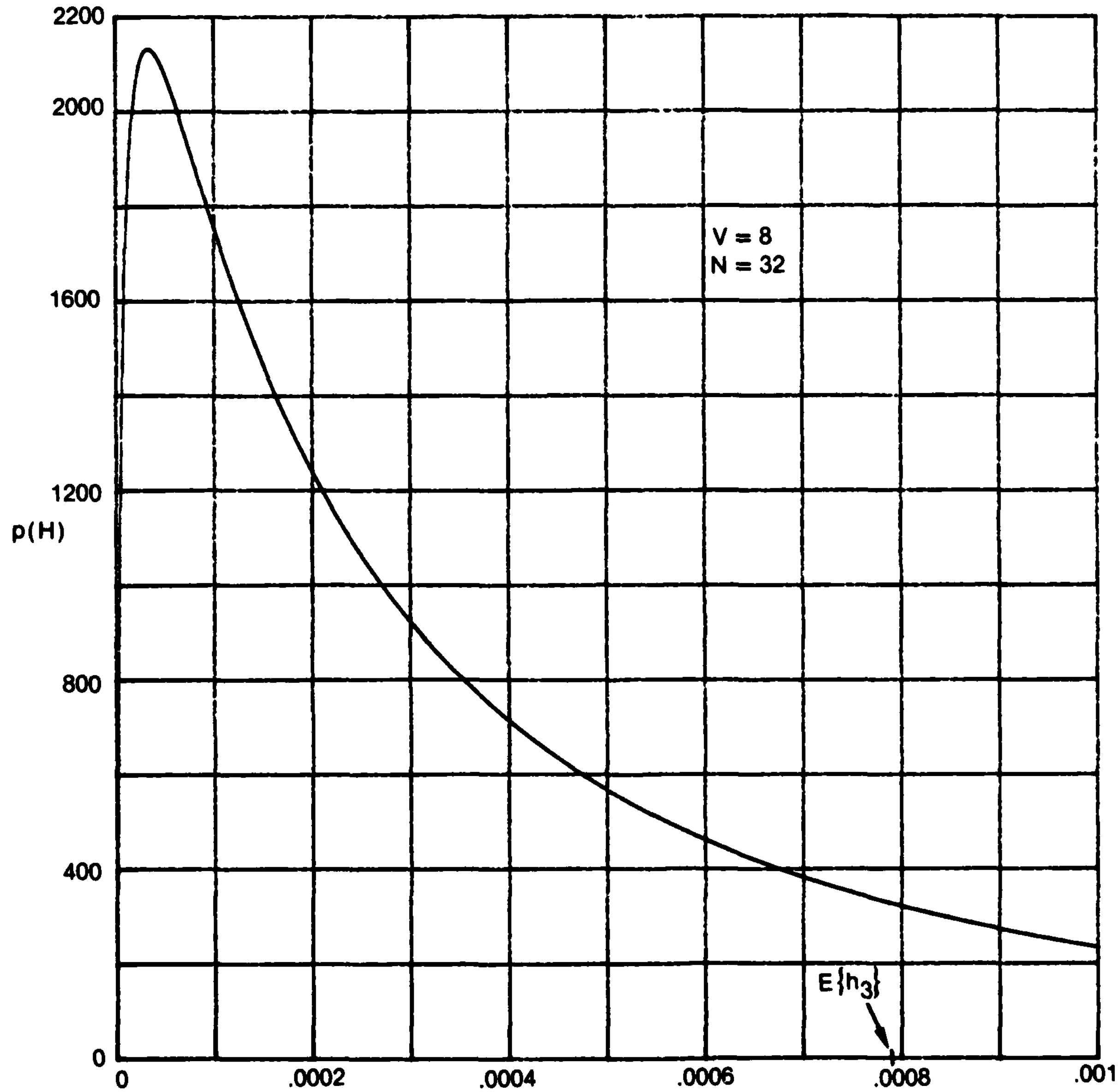
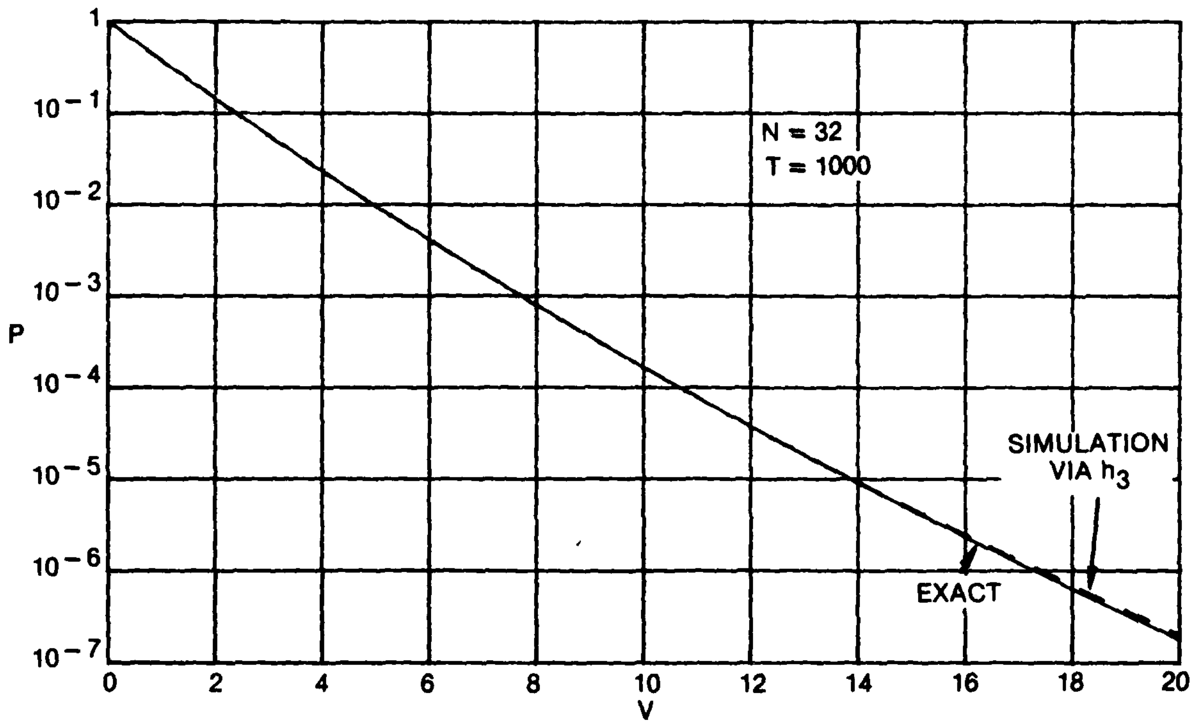
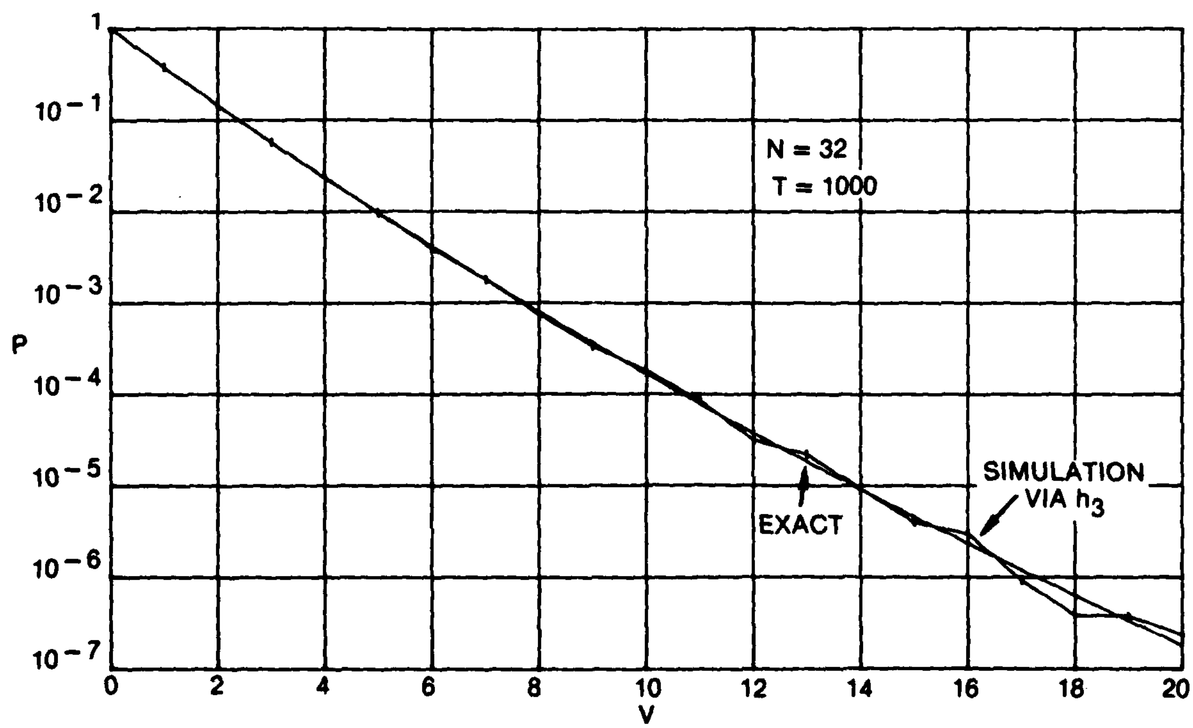


Figure 7. Probability Density Function for h_3



(a) SAME RANDOM NUMBERS FOR EACH V



(b) DIFFERENT RANDOM NUMBERS FOR EACH V

Figure 8. Simulation for a Shifted PDF

solid (exact) curve in figures 8(a) and 8(b) would not be available, the dashed curve in figure 8(a) would give no indication of how reliable the result was, whereas the fluctuating result in 8(b) would give a rough idea of the reliability of the estimate, since each plotted point is independent of its neighbor. The "in-breeding" of the same data in figure 8(a) saves time but can be a dangerous and misleading procedure. A program for the simulation result of figure 8(b) is given in appendix F.

A measure of the stability of the results in figure 8 is afforded by the variance of α_3. To determine this quantity, we first need ν-th moment

$$E\left\{h_3^{\nu}(X)\right\} = E\left\{\exp\left(-\frac{V}{N}\,s\nu\right)\right\} = \int_0^{\infty} ds\, \frac{s^{N-1}\,e^{-s}}{(N-1)!}\exp\left(-\frac{V}{N}\,s\nu\right)$$
$$= \left(1 + \nu\,\frac{V}{N}\right)^{-N} , \tag{46}$$

where we have used (42) and (41). Then the variance of h_3 is

$$V = \left(1 + 2\frac{V}{N}\right)^{-N} - \left(1 + \frac{V}{N}\right)^{-2N} , \tag{47}$$

and that for α_3 is T times smaller, for T independent trials. A table of the mean and standard deviation of α_3 follows below. These standard deviations are 3-4 times smaller than those given in table 1, which were for the optimum scaling.

Table 2. Statistics of α_3 for N = 32, T = 1000

V	P = E$\{\alpha_3\}$	SD$\{\alpha_3\}$
2	1.44E-1	1.56E-3
4	2.31E-2	5.10E-4
6	4.09E-3	1.44E-4
8	7.92E-4	4.11E-5
10	1.66E-4	1.23E-5
12	3.76E-5	3.91E-6
14	9.05E-6	1.32E-6
16	2.32E-6	4.77E-7
18	6.28E-7	1.82E-7
20	1.79E-7	7.31E-8

A Gated Conditional PDF

The result in the previous subsection was obtained by modifying conditional PDF $p(x_{N+1}|s)$; here we take the opposite tack by modifying $p(s|x_{N+1})$. First define a gate function

$$U_s(a, b) = \left\{\begin{matrix} 1 \text{ for } a < s < b \\ \\ 0 \text{ otherwise} \end{matrix}\right\} . \tag{48}$$

Then define alternative PDF

$$p^*\left(s,\ x_{N+1}\right) = p^*\left(x_{N+1}\right) \cdot p^*\left(s \mid x_{N+1}\right)$$

$$= e^{-x_{N+1}} \cdot \frac{s^{N-1}\ e^{-s}}{(N-1)!}\ U_s\left(0,\ \frac{N}{V}\ x_{N+1}\right) \Big/ D_N\left(\frac{N}{V}\ x_{N+1}\right) \quad \text{for } x_{N+1} > 0, \tag{49}$$

where denominator D_N must be determined so that the conditional PDF has unit volume; that is, by use of (48) and (32),

$$D_N\left(\frac{N}{V}\ x_{N+1}\right) = \int_0^{\frac{N}{V} x_{N+1}} ds\ \frac{s^{N-1}\ e^{-s}}{(N-1)!}$$

$$= 1 - \exp\left(-\ \frac{N}{V}\ x_{N+1}\right)\ e_{N-1}\left(\frac{N}{V}\ x_{N+1}\right) \quad \text{for } x_{N+1} > 0. \tag{50}$$

The unit gated function U_s in (49) keeps $p^* > 0$ only in the region R_s where $x_{N+1} > \frac{V}{N}s$, as was indicated desirable in the previous section. The use of (17), (39), (49), and (50) leads to counting function

$$h_4(X) = h_4\left(s,\ x_{N+1}\right) = D_N\left(\frac{N}{V}\ x_{N+1}\right) \quad \text{for } x_{N+1} > 0 \quad . \tag{51}$$

Since random variable s is not used in (51), there is no need to generate it; we use (51) with the PDF $p^*(x_{N+1}) = \exp(-x_{N+1})$ for $x_{N+1} > 0$.

The exceedance probability of h_4 may be found as follows:

$$\text{Prob}\{h_4 > H\} = \text{Prob}\left\{D_N\left(\frac{N}{V}\ x_{N+1}\right) > H\right\} = \text{Prob}\left\{x_{N+1} > \frac{V}{N}\ \tilde{D}_N(H)\right\}$$

$$= \int_{\frac{V}{N}\tilde{D}_N(H)}^{\infty} dx_{N+1}\ \exp\left(-\ x_{N+1}\right) = \exp\left(-\ \frac{V}{N}\ \tilde{D}_N(H)\right) \quad \text{for } 0 < H < 1 \quad , \tag{52}$$

where $\tilde{D}_N$ is the inverse function to D_N, i.e.,

$$\tilde{D}_N(D_N(y)) = y \quad . \tag{53}$$

The PDF of h_4 is available through differentiation with respect to H:

$$p(H) = \frac{V}{N}\ \tilde{D}_N'(H)\ \exp\left(-\frac{V}{N}\ \tilde{D}_N(H)\right) = \frac{V}{N}\ \frac{\exp\left(-\frac{V}{N}\ \tilde{D}_N(H)\right)}{D_N'\left(\tilde{D}_N(H)\right)}$$

$$= \frac{V}{N}\ \frac{\exp\left[\left(1 - \frac{V}{N}\right)\tilde{D}_N(H)\right]\ (N-1)!}{\left[\tilde{D}_N(H)\right]^{N-1}} \quad \text{for } 0 < H < 1 \quad . \tag{54}$$

Here we used the result of differentiating (53) with respect to y and the derivative of (50), namely,

$$\tilde{D}_N'\left(D_N(y)\right)\ D_N'(y) = 1 \quad ,$$

$$D_N'(y) = \frac{y^{N-1}\ e^{-y}}{(N-1)!} \quad \text{for } y > 0 \quad . \tag{55}$$

The numerical calculation of (52) and (54) can be achieved without the need of calculating the inverse function $\tilde{D}_N(H)$. We employ a parametric approach by choosing a value for $a = \tilde{D}_N(H)$; then from (52) through (54), we can compute

$$H = D_N(a),\ \mathrm{Prob}\{h_4 > H\} = \exp\left(-\frac{V}{N}a\right) \ ,\ p(H) = \frac{V}{N}\ \frac{\exp\left[\left(1 - \frac{V}{N}\right)a\right]}{a^{N-1}} \ , \tag{56}$$

all in terms of the parameter a. The function

$$D_N(a) = 1 - \exp(-a)\ \ e_{N-1}(a) \tag{57}$$

defined in (50) and (32) must, of course, still be evaluated.

The exceedance probability (52) and PDF (54) are presented in figure 9 for $V = 8$, $N = 32$. There is a large undesirable cusp in the PDF at $H = 0+$, and a lesser one at $H = 1-$. This choice of alternative PDF in (49) gives results reminiscent of the PDF for h_1 in the direct simulation, and is not expected to be very useful. A simulation result in figure 10 confirms this. The simulation run in figure 10a employed the same random numbers at all V, for each of the 1000 trials. Although a very smooth estimation curve results in figure 10a, it is totally misleading; for example, it indicates probabilities at $V = 14$ which are two orders of magnitude too small. If the exact answer were not available, which is the practical situation, the smoothness of the estimate might give a false sense of reliability; in reality, the smoothness of the estimated curve is no measure of the accuracy of the result when the data are so strongly inbred by being used repeatedly. For contrast, the simulation in figure 10b was run with different random numbers for all V, for each of the 1000 trials. The extremely large fluctuations in the estimates for the lower values of probability are indicative of the unreliability of this importance sampling procedure.

The variance of h_4 can be evaluated as follows from (51) and (50):

$$h_4 = \int_0^{rx_{N+1}} ds\ \frac{s^{N-1}\ e^{-s}}{(N-1)!} \tag{58}$$

where $r \equiv N/V$. Then using (49), we obtain the mean value as

$$E\{h_4\} = \int_0^{\infty} dx\ e^{-x} \int_0^{rx} ds\ \frac{s^{N-1}\ e^{-s}}{(N-1)!}$$

$$= \int_0^{\infty} ds\ \frac{s^{N-1}\ e^{-s}}{(N-1)!} \int_{s/r}^{\infty} dx\ e^{-x} = \left(1 + \frac{V}{N}\right)^{-N} \ , \tag{59}$$

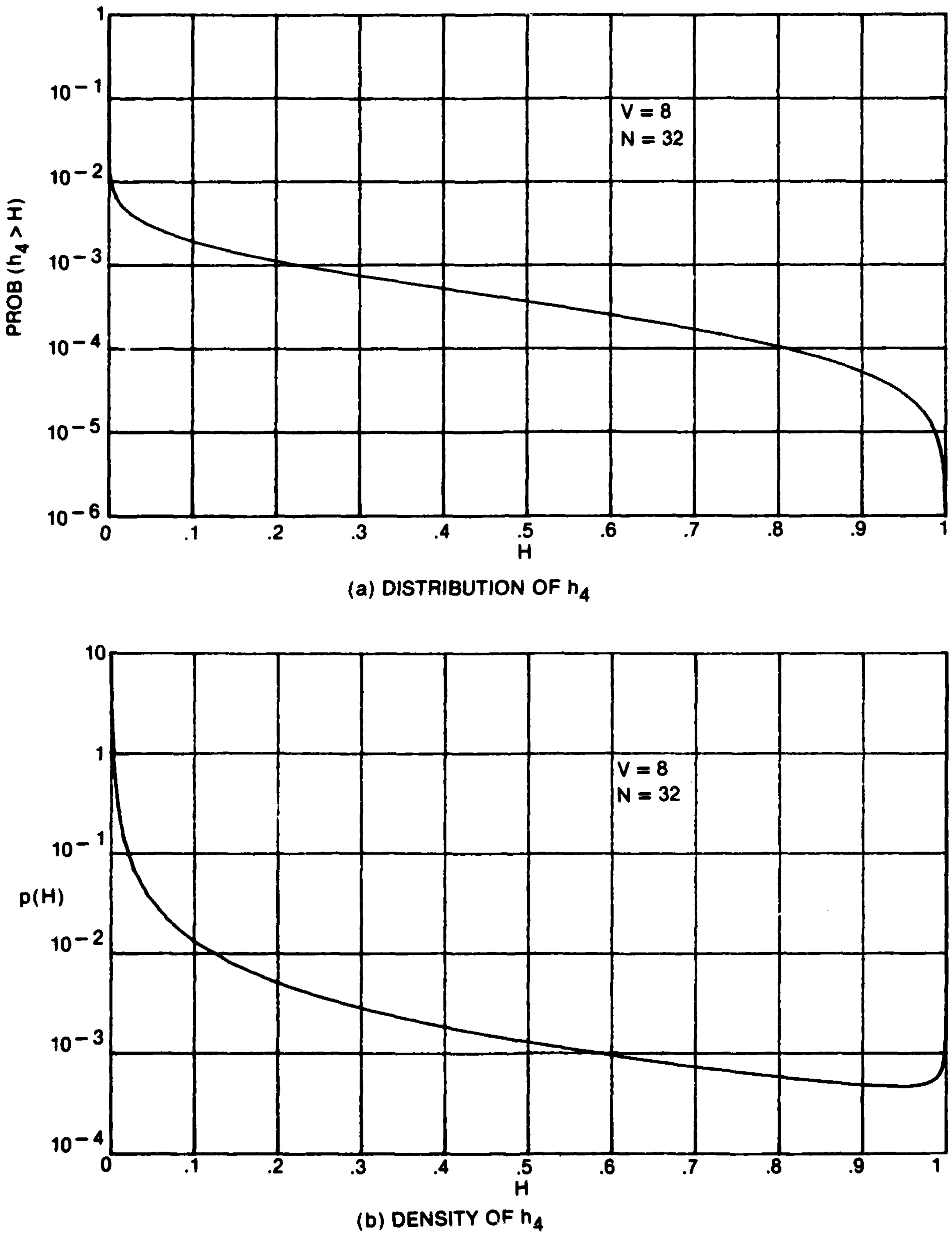
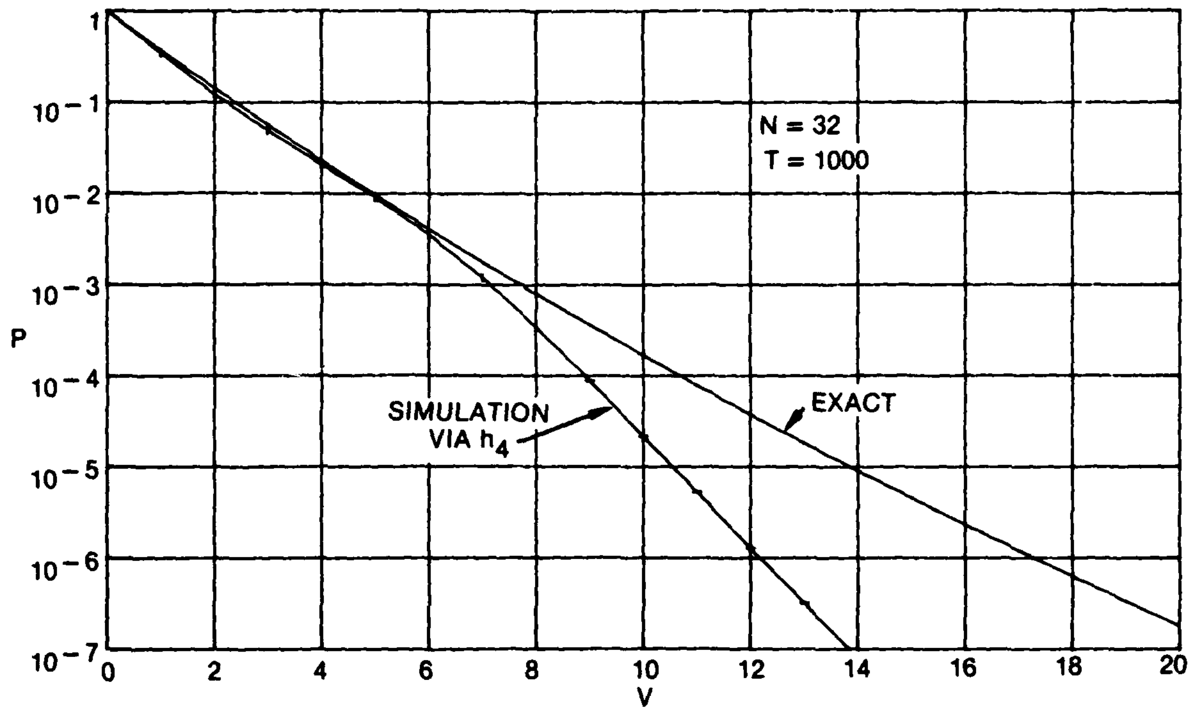
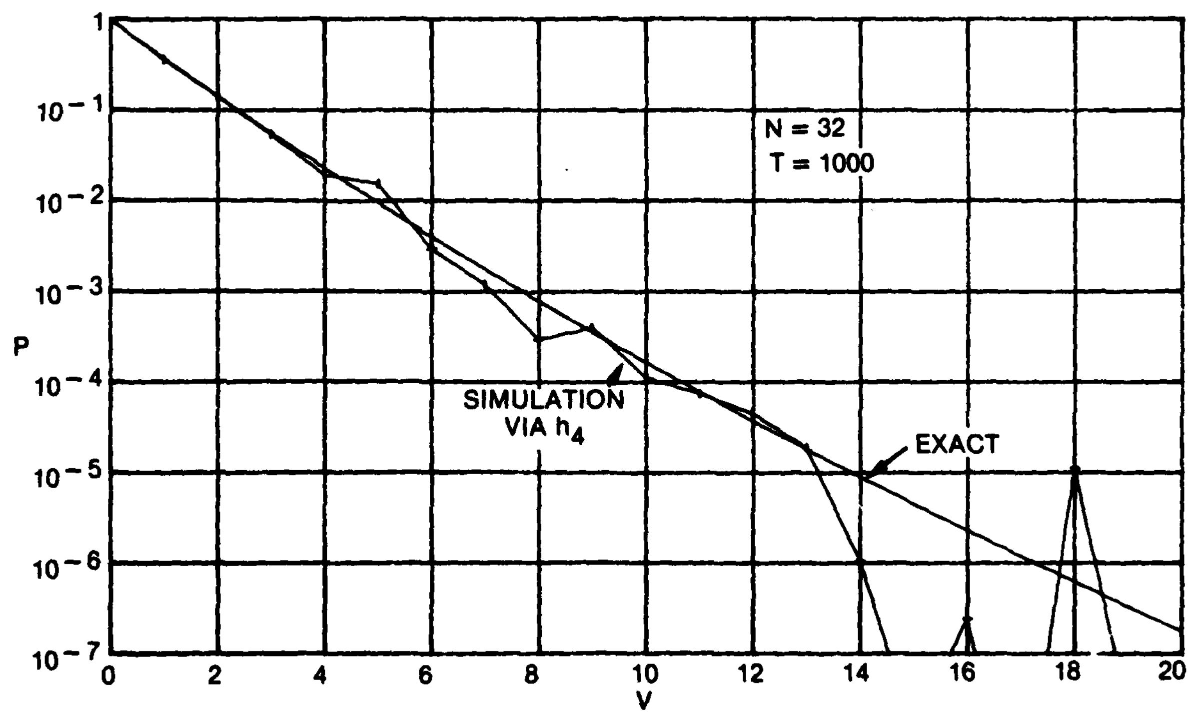


Figure 9. Distribution and Density Functions for h_4



(a) SAME RANDOM NUMBERS FOR EACH V



(b) DIFFERENT RANDOM NUMBERS FOR EACH V

Figure 10. Simulation for a Gated Conditional PDF

in agreement with (5) as expected. Also, letting $p_s(\,)$ denote the PDF of s, and $P_s(\,)$ its cumulative distribution, we have

$$E\{h_4^2\} = \int_0^\infty dx\ e^{-x} \int_0^{rx}\!\!\int ds\ dt\ p_s(s)\ p_s(t)$$

$$= \int_0^\infty dx\ e^{-x}\ 2\int_0^{rx} ds\ p_s(s) \int_0^{s} dt\ p_s(t) = 2\int_0^\infty dx\ e^{-x} \int_0^{rx} ds\ p_s(s)\ P_s(s)$$

$$= 2\int_0^\infty ds\ p_s(s)\ P_s(s) \int_{s/r}^\infty dx\ e^{-x} = 2\int_0^\infty ds\ \frac{s^{N-1}\ e^{-s}}{(N-1)!}\ e^{-s/r} \int_0^{s} dt\ \frac{t^{N-1}\ e^{-t}}{(N-1)!}$$

$$= 2\int_0^\infty ds\ \frac{s^{N-1}\ e^{-sq}}{(N-1)!} \left[1 - e^{-s} \sum_{n=0}^{N-1} \frac{1}{n!}\ s^n\right]$$

$$= 2\left[\frac{1}{q^N} - \sum_{n=0}^{N-1} \binom{N-1+n}{n} \frac{1}{(1+q)^{N+n}}\right]$$

$$=2\left(1+\frac{V}{N}\right)^{-N} - 2\left(2+\frac{V}{N}\right)^{-N} \sum_{n=0}^{N-1} \binom{N-1+n}{n} \left(2+\frac{V}{N}\right)^{-n}\ , \tag{60}$$

where we temporarily let $q = 1+1/r = 1 + V/N$. The variance of h_4 is equal to (60) minus the square of (59).

The mean and standard deviation of

$$\alpha_4 = \frac{1}{T} \sum_{i=1}^{T} h_4(X^{(i)}) \tag{61}$$

are given in table 3 for $N = 32$, $T = 1000$. Comparison with tables 1 and 2 for α_2 and α_3, respectively, reveals that the performance results in table 3 are much poorer. In fact, the results for $SD\{\alpha_4\}$ are only 2-3 times better than for the direct simulation case α_1; this is in keeping with the observation made under (57) regarding the PDF of h_4 in figure 9.

Table 3. Statistics of α_4 for N = 32, T = 1000

V	P = E{α_4}	SD{α_4}
2	1.44E-1	9.78E-3
4	2.31E-2	3.77E-3
6	4.09E-3	1.42E-3
8	7.92E-4	5.44E-4
10	1.66E-4	2.15E-4
12	3.75E-5	8.78E-5
14	9.05E-6	3.68E-5
16	2.32E-6	1.58E-5
18	6.28E-7	6.93E-6
20	1.79E-7	3.11E-6

A Combined Scaled and Shifted PDF

Since counting functions h_2 and h_3 performed rather well, an attempt at combining their features was attempted. Instead of the alternative conditional PDF considered in (41), we tried

$$p^*(x_{N+1}|s) = \frac{1}{K} \exp\left[-\frac{1}{K}\left(x_{N+1} - \frac{V}{N}s\right)\right] \quad \text{for } x_{N+1} > \frac{V}{N}s,\ K > 1 \ . \tag{62}$$

The counting function is now a generalization of (42):

$$h_5 = K \exp\left[-x_{N+1}\left(1 - \frac{1}{K}\right) - \frac{V}{KN}s\right] \quad \text{for } x_{N+1} > \frac{V}{N}s > 0 \ . \tag{63}$$

The ν-th moment of h_5 is given by

$$E\{h_5^\nu\} = \int_0^\infty ds \int_0^\infty dx\ p^*(s, x)\ h_5^\nu$$

$$= \int_0^\infty ds \frac{s^{N-1}e^{-s}}{(N-1)!} \int_{sV/N}^\infty dx \frac{1}{K} \exp\left[-\frac{x - sV/N}{K}\right] K^\nu \exp\left[-\nu x \frac{K-1}{K} - \nu\frac{V}{KN}s\right]$$

$$= \frac{K^\nu}{1 + \nu(K-1)} \int_0^\infty ds \frac{s^{N-1}}{(N-1)!} \exp\left[-s(1 + \nu V/N)\right]$$

$$= \frac{K^\nu}{\left[1 + \nu(K-1)\right]\left(1 + \nu\frac{V}{N}\right)^N} \tag{64}$$

when the denominator terms are positive. For $\nu = 1$, this equals (5) as it should, independently of K. For $K \geq 1$, (64) is minimized by the choice of scaling $K = 1$, regardless of the values of V, N, and $\nu(>1)$. Thus the minimum variance of h_5 is attained by not scaling at all, and just using the shifted PDF, as done with h_3. Accordingly this alternative PDF was not studied any further.

CONCLUSIONS

The importance sampling procedure is an important and useful tool for estimating small probabilities. Not only can it estimate probabilities considerably less than 1/T, where T is the number of independent trials, but it can do so with arbitrarily small variance.

However, the major flaw is that the exact alternative PDF to use for data generation is not known. Some guidelines for choosing good PDFs have been derived. They indicate that the new PDF should mimic the given PDF in the region where the original PDF is positive and where the test under consideration yields threshold crossings. In fact, one should use a PDF which never generates data that lead to processor outputs less than the threshold value(s) under investigation. The difficulty of satisfying these goals makes selection of an alternative PDF more of an art than a science. Several procedures were investigated here, and at least one gave remarkably good estimations of probabilities in the 10^{-7} range, by means of only 1000 trials. Some other choices yielded poorer results. It may be necessary to try several different guesses for the alternative PDF, and then select the best.

The danger of being deceived by a smooth estimation curve, of the exceedance probability versus threshold, is great if one employs the same data for all the threshold values considered. Rather, it is recommended that different random numbers be used for each threshold considered. Then the width of the independent fluctuations at different thresholds serves as a measure of the reliability of the results obtained. Of course, this additional feature is achieved at the expense of more computer processing time, since new data must be generated each time the threshold is changed.

Since the region of data space where the threshold is exceeded depends on the threshold value itself, it may be necessary to make different choices of the alternative PDF for each threshold value of interest. This drawback is one of the compensating features that must be accepted for the ability to estimate small probabilities with vanishingly small error. Importance sampling is not a panacea.

Appendix A

GENERALIZED LIKELIHOOD RATIO

The PDF for noise-only is given by (2). For a given observation X, (2) is maximized by the choice of β as

$$\beta_o = \frac{1}{N+1} \sum_{n=1}^{N+1} x_n \quad . \tag{A-1}$$

The corresponding maximum value of (2) is

$$\hat{p}_o(X) = \frac{\exp(-N-1)}{\beta_o^{N+1}} \quad . \tag{A-2}$$

The PDF for signal-plus-noise is given by (3); it is maximized by the choices

$$\beta_1 = \frac{1}{N} \sum_{n=1}^{N} x_n \ , \quad \gamma_1 = x_{N+1} \quad , \tag{A-3}$$

provided that $\gamma_1 \geqslant \beta_1$. If $x_{N+1} < \frac{1}{N} \sum_{n=1}^{N} x_n$, then we cannot accept γ_1 and β_1 as given by (A-3), because then we would have $\gamma_1 < \beta_1$, which is inconsistent with the precondition stated with (3) that $\gamma > \beta$. Instead we would set $\gamma = \beta$ and maximize (3), getting

$$\hat{\gamma}_1 = \hat{\beta}_1 = \frac{1}{N+1} \sum_{n=1}^{N+1} x_n = \beta_o \quad \text{if} \quad x_{N+1} < \frac{1}{N} \sum_{n=1}^{N} x_n \quad . \tag{A-4}$$

Thus the maximum value of (3) is given by

$$\hat{p}_1(X) = \left\{ \begin{array}{ll} \dfrac{\exp(-N-1)}{\beta_1^N \, x_{N+1}} & \text{for} \quad x_{N+1} \geq \dfrac{1}{N} \displaystyle\sum_{n=1}^{N} x_n \\ \dfrac{\exp(-N-1)}{\beta_o^{N+1}} & \text{for} \quad x_{N+1} < \dfrac{1}{N} \displaystyle\sum_{n=1}^{N} x_n \end{array} \right\} \quad . \tag{A-5}$$

The generalized likelihood ratio is given by the ratio of (A-5) to (A-2):

$$\text{GLR} = \frac{\beta_o^{N+1}}{\beta_1^N \, x_{N+1}} = \frac{N^N}{(N+1)^{N+1}} \, \frac{(1+r)^{N+1}}{r} \quad \text{for} \quad r \geq \frac{1}{N} \quad , \tag{A-6}$$

and GLR = 1 for $r < 1/N$, where

$$r \equiv \frac{x_{N+1}}{x_1 + x_2 + \ldots + x_N} \quad . \tag{A-7}$$

But the generalized likelihood ratio in (A-6) is a monotonically increasing function of r for $r \geq 1/N$. Therefore the generalized likelihood ratio test is equivalent to comparing r with a threshold; i.e.,using (A-7), the detection statistic is

$$\frac{x_{N+1}}{\frac{1}{N}\left(x_1 + x_2 \ldots + x_N\right)} \underset{H_o}{\overset{H_1}{\gtrless}} V \ , \tag{A-8}$$

where threshold $V \geq 1$.

In order to evaluate the false alarm probability of test (A-8), we let

$$s = x_1 + x_2 + \ldots + x_N \ . \tag{A-9}$$

Then from (2), the PDF of s is

$$p(s) = \frac{s^{N-1} \ e^{-s}}{(N - 1)!} \quad \text{for} \quad s > 0 \tag{A-10}$$

where we have let $\beta = 1$, since absolute scale is irrelevant to test (A-8). Then

$$P_{FA} = \text{Prob}\left\{x_{N+1} > s\frac{V}{N} \middle| H_o\right\}$$

$$= \int_0^\infty ds \ \frac{s^{N-1} e^{-s}}{(N - 1)!} \int_{sV/N}^\infty dx \ e^{-x} = \frac{1}{(1 + V/N)^N} \ . \tag{A-11}$$

Appendix B

PROGRAM FOR SCALING OF POTENTIAL-SIGNAL SAMPLE

```
10   N=32
20   T=1000
30   K=6
40   DIM A(20)
50   K1=1-1/K
60   Random=SQR(.6)
70   RANDOMIZE Random
80   FOR I=1 TO T
90   X=RND
100  FOR J=2 TO N
110  X=X*RND
120  NEXT J
130  S=-LOG(X)            ! EQ 25
140  X=-K*LOG(RND)        ! EQ 23
150  E=EXP(-X*K1)
160  Vc=INT(N*X/S)
170  Vc=MIN(Vc,20)
180  FOR V=0 TO Vc
190  A(V)=A(V)+E
200  NEXT V
210  NEXT I
220  R=K/T
230  FOR V=0 TO 20
240  A(V)=LGT(A(V)*R)
250  NEXT V
260  PLOTTER IS "GRAPHICS"
270  GRAPHICS
280  SCALE 0,20,-7,0
290  GRID 2,1
300  PENUP
310  LINE TYPE 9
320  FOR V=0 TO 20
330  PLOT V,A(V)      ! SIMULATION
340  NEXT V
350  PENUP
360  LINE TYPE 1
370  FOR V=0 TO 20
380  PLOT V,-N*LGT(1+V/N) ! EXACT
390  NEXT V
400  PENUP
410  END
```

Appendix C

MOMENTS OF h_2

Counting function h_2 is given by (24), where s is given by (25). By reference to (22), it can be seen that the PDF of s is

$$p(s) = \frac{s^{N-1}\, e^{-s}}{(N-1)!} \quad \text{for } s > 0 \quad , \tag{C-1}$$

while that for x_{N+1} is

$$p(x_{N+1}) = \exp(-x_{N+1}) \quad \text{for } x_{N+1} > 0 \quad . \tag{C-2}$$

Since only the PDF of random variable x_{N+1} is changed in alternative PDF p* in (23), we have

$$p^*(s,\, x_{N+1}) = \frac{s^{N-1}\, e^{-s}}{(N-1)!}\, \frac{1}{K} \exp\left(-\frac{x_{N+1}}{K}\right) \quad \text{for} \quad s > 0 \;,\; x_{N+1} > 0 \quad . \tag{C-3}$$

Since h_2 in (24) is non-zero only if $x_{N+1} > sV/N$, then the ν-th moment of h_2 is given by

$$E\left\{h_2^\nu\right\} = \int_0^\infty ds\, \frac{s^{N-1}\, e^{-s}}{(N-1)!} \int_{sV/N}^\infty dx\, \frac{\exp(-x/K)}{K}\, K^\nu \exp\left(-\nu x\left(1 - \frac{1}{K}\right)\right)$$

$$= \frac{K^\nu}{1 + \nu(K-1)} \int_0^\infty ds\, \frac{s^{N-1}}{(N-1)!} \exp\left[-s\left\{1 + \frac{V}{N}\left(\frac{1}{K} + \nu - \frac{\nu}{K}\right)\right\}\right]$$

$$= \frac{K^{\nu-1}}{\nu - \frac{\nu-1}{K}}\, \frac{1}{\left[1 + \frac{V}{N}\left(\nu - \frac{\nu-1}{K}\right)\right]^N} \quad . \tag{C-4}$$

For $\nu = 1$, this reduces to (28).

The mean square value of h_2 is given by substituting $\nu = 2$ in (C-4):

$$E\left\{h_2^2\right\} = \frac{K}{2 - \frac{1}{K}}\, \frac{1}{\left[1 + \frac{V}{N}\left(2 - \frac{1}{K}\right)\right]^N} \quad . \tag{C-5}$$

We want to minimize this expression by choice of scaling K. To do this, let $t = 1/K$, and consider the reciprocal of (C-5):

$$R = t(2-t)(a - bt)^N \;, \quad \text{where} \quad a \equiv 1 + \frac{2V}{N} \;,\; b \equiv \frac{V}{N} \quad . \tag{C-6}$$

Setting dR/dt to zero, we must solve the equation

$$(2 - t)(a - bt) - t(a - bt) - Nbt(2 - t) = 0 \quad . \tag{C-7}$$

If we simplify and put $t = 1/K_o$, there follows

$$2a\, K_o^2 - 2(a + b + Nb)\, K_o + (N + 2)b = 0 \quad . \tag{C-8}$$

Solving this quadratic, and substituting the values for a and b in (C-6), we find for the optimum value of scaling,

$$K_o = \frac{1 + V + \frac{3V}{N} + \left(1 + \frac{2V}{N} + \left(V + \frac{V}{N}\right)^2\right)^{1/2}}{2\left(1 + \frac{2V}{N}\right)} \quad . \tag{C-9}$$

The negative square root is discarded because it leads to values of $K_o < 1$, which are disallowed.

Appendix D

DISTRIBUTION AND DENSITY OF h_2

Distribution

We repeat from (24)

$$h_2 = K \exp\left(-x_{N+1} \frac{K-1}{K}\right) U\left(x_{N+1} - \frac{V}{N} s\right) . \tag{D-1}$$

Now $h_2 = H$ when $x_{N+1} = a$, where

$$K \exp\left(-a \frac{K-1}{K}\right) = H ; \quad a = \ln\left(\frac{K}{H}\right) \frac{K}{K-1} . \tag{D-2}$$

Also $h_2 > H$ when x_{N+1} lies in region R_a in figure D-1.

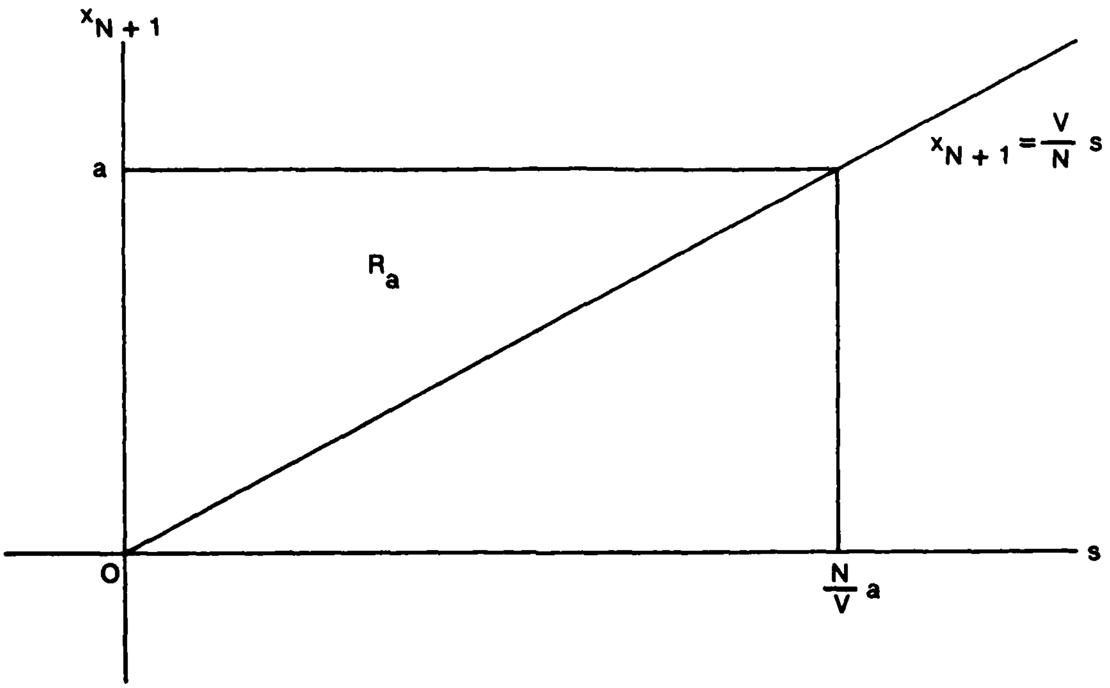


Figure D-1. Region R_a where $h_2 > H$

Therefore, using (C-3) and figure D-1, we have

$$\mathrm{Prob}\{h_2 > H\} = \int_0^{aN/V} ds \frac{s^{N-1} e^{-s}}{(N-1)!} \int_{sV/N}^{a} dx \frac{1}{K} \exp\left(-\frac{x}{K}\right)$$

$$= \int_0^{aN/V} ds \frac{s^{N-1} e^{-s}}{(N-1)!} \left[\exp\left(-\frac{V}{NK} s\right) - \exp\left(-\frac{a}{K}\right)\right]$$

$$= \left(1 + \frac{V}{KN}\right)^{-N} \left[1 - e^{-A_1} e_{N-1}(A_1)\right]$$

$$- \exp\left(-\frac{a}{K}\right)\left[1 - e^{-A_2} e_{N-1}(A_2)\right] \quad , \tag{D-3}$$

where a is given in (D-2),

$$A_1 \equiv a\left(\frac{N}{V} + \frac{1}{K}\right) , \quad A_2 \equiv a \frac{N}{V} \quad , \tag{D-4}$$

and (reference 3, eq. 6.5.11)

$$e_M(x) \equiv \sum_{m=0}^{M} \frac{1}{m!} x^m \tag{D-5}$$

is the leading terms, through x^M, of the power series expansion of e^x.

As $H \to 0+$, $a \to +\infty$ from (D-2). Then $A_1 \to +\infty$ and $A_2 \to +\infty$ from (D-4), and $e_{N-1}(A_j) \sim A_j^{N-1}/(N-1)!$. However, the exponential $\exp(-A_j)$ dominates this latter behavior, and (D-3) yields

$$\mathrm{Prob}\{h_2 > 0\} = \left(1 + \frac{V}{KN}\right)^{-N} \quad . \tag{D-6}$$

We see, directly from (D-1), that h_2 can never exceed K. When we substitute $H = K$ in (D-2), we get $a = 0$, and (D-3)-(D-5) then yield $\mathrm{Prob}\{h_2 > K\} = 0$, as expected.

Density

An alternative way of expressing (D-3) is as follows; by reference to figure D-1,

$$\mathrm{Prob}\{h_2 > H\} = \int_0^{a} dx \frac{1}{K} \exp\left(-\frac{x}{K}\right) \int_0^{xN/V} ds \frac{s^{N-1} e^{-s}}{(N-1)!}$$

$$= \int_0^{a} dx \frac{1}{K} \exp\left(-\frac{x}{K}\right)\left[1 - \exp\left(-\frac{N}{V} x\right) e_{N-1}\left(\frac{N}{V} x\right)\right] , \tag{D-7}$$

where the integrand of (D-7) is independent of H. But by definition,

$$\mathrm{Prob}\{h_2 > H\} = \int_H^{\infty} dh_2 \; p(h_2) \quad , \tag{D-8}$$

where $p(h_2)$ is the PDF of h_2. Setting the right-hand sides of (D-7) and (D-8) equal to each other, and differentiating with respect to H, we obtain

$$-p(H) = \frac{\partial a}{\partial H} \frac{1}{K} \exp\left(-\frac{a}{K}\right)\left[1 - \exp\left(-\frac{N}{V}a\right) e_{N-1}\left(\frac{N}{V}a\right)\right] . \tag{D-9}$$

But from (D-2),

$$\frac{\partial a}{\partial H} = -\frac{K}{K-1}\frac{1}{H} . \tag{D-10}$$

Therefore the PDF of h_2 is given by

$$p(H) = \frac{\exp(-a/K)}{(K-1)H}\left[1 - \exp\left(-\frac{N}{V}a\right) e_{N-1}\left(\frac{N}{V}a\right)\right] \quad \text{for} \quad 0 < H \le K . \tag{D-11}$$

As $H \to 0+$, the bracketed term in (D-11) tends to 1 since $a \to +\infty$. Therefore,

$$p(H) \sim \left((K-1)\, K^{\frac{1}{K-1}}\, H^{\frac{K-2}{K-1}}\right)^{-1} \quad \text{as} \quad H \to 0+ . \tag{D-12}$$

This infinite cusp at the origin is integrable. For the simulation result in figure 5 for $K = 6$, this yields $p(H) \sim .14/H^{.8}$ as $H \to 0+$.

Appendix E

DERIVATION OF OPTIMUM DENSITY FOR p*

It is convenient to define three regions in X-space; namely,

$$\left.\begin{aligned} R_V &: g(X) > V \\ R &: p(X) > 0 \\ R^* &: p^*(X) > 0 \end{aligned}\right\} \quad . \tag{E-1}$$

Now we define counting function (more precisely than (17)) as

$$h(X) = \left\{ \begin{array}{l} \frac{p(X)}{p^*(X)} U(g(X) - V) \quad \text{for} \quad X\varepsilon R^* \\ 0 \text{ otherwise} \end{array} \right\} \quad . \tag{E-2}$$

Then the mean value of h is obtained by averaging over p*:

$$E\{h(X)\} = \int_{R^*} dX\ p^*(X)\ h(X) = \int_{R^*} dX\ p(X)\ U(g(X) - V) \quad . \tag{E-3}$$

The integrand of (E-3) is non-zero in region $R \cap R_V$. In order to keep h unbiased, we henceforth assume that $R^* \supset (R \cap R_V)$; for then $E\{h(X)\} = P$, according to (8).

According to (20) and (21), we now want to minimize

$$E\{h^2(X)\} = \int_{R^*} dX\ p^*(X)\ h^2(X) = \int_{R^*} dX\ \frac{p^2(X)}{p^*(X)} U(g(X) - V) \tag{E-4}$$

by choice of p*(X). If we let

$$A(X) \equiv p^2(X)\ U(g(X) - V) \quad \text{for} \quad X\varepsilon R^* \quad , \tag{E-5}$$

then (E-4) can be expressed as

$$E\{h^2\} = \int_{R^*} dX\ \frac{A(X)}{p^*(X)} \quad . \tag{E-6}$$

A(X) combines all the given known quantities in one expression.

We have the constraint that the volume under p* must be unity for a legal PDF. If we let $p_o(X)$ be the optimum value of $p^*(X)$, and perform a perturbation $\epsilon\eta(X)$ of $p_o(X)$, using a Lagrange multiplier for the constraint, the perturbed value of (E-6) becomes

$$\int_{R^*} dX \frac{A(X)}{p_o(X) + \epsilon\eta(X)} - \lambda \int_{R^*} dX \left[p_o(X) + \epsilon\eta(X)\right] \quad . \tag{E-7}$$

Differentiating with respect to ϵ, and setting $\epsilon = 0$, we must obtain a zero quantity for all variations $\eta(X)$, in order for $p_o(X)$ to be the optimum. There follows for the optimum PDF

$$p_o(X) = c(A(X))^{1/2} = c\ p(X)\ U(g(X) - V) \quad \text{for } X\epsilon R^* \quad , \tag{E-8}$$

where c is a positive constant and we used (E-5). The right-hand side of (E-8) is non-negative, as it must be for a legal PDF. An alternative statement of (E-8) is obviously

$$p_o(X) = \begin{Bmatrix} c\ p(X) & \text{for } X\epsilon(R \cap R_V) \\ & \\ 0 \text{ otherwise} & \end{Bmatrix} \quad . \tag{E-9}$$

The constant in (E-9) is determined by satisfying the constraint of unit volume for a PDF:

$$1 = \int_{R^*} dX\ p_o(X) = c \int_{R \cap R_V} dX\ p(X) = c\ P \quad , \tag{E-10}$$

using (8). Thus $c = 1/P$, giving for the optimum PDF

$$p_o(X) = \begin{Bmatrix} p(X)/P & \text{for } X\epsilon(R \cap R_V) \\ & \\ 0 \text{ otherwise} & \end{Bmatrix} \quad . \tag{E-11}$$

The minimum mean square value of h follows from (E-4) as

$$\begin{aligned} \min E\{h^2(X)\} &= P \int_{R^*} dX\ p(X)\ U(g(X) - V) \\ &= P \int_{R \cap R_V} dX\ p(X) = P^2 \quad . \end{aligned} \tag{E-12}$$

There follows for the variance of the optimum h, namely h_o,

$$V\{h_o\} = E\{h_o^2\} - E^2\{h_o\} = P^2 - P^2 = 0 \quad . \tag{E-13}$$

Use of (E-11) in (E-2) shows that the optimum counting function is

$$h_o(X) = \begin{Bmatrix} P & \text{for} \quad X\varepsilon(R \cap R_v) \\ \\ 0 & \text{otherwise} \end{Bmatrix} . \tag{E-14}$$

That is, every trial generated according to optimum PDF $p_o(X)$ yields the same value for h_o, namely P. The value 0 is never generated because $p_o(X)$ is zero for such data values X.

Appendix F

PROGRAM FOR A SHIFTED PDF

```
10    N=32
20    T=1000
30    DIM A(20)
40    Random=SQR(.6)
50    RANDOMIZE Random
60    FOR I=1 TO T
70    FOR V=0 TO 20
80    X=RND
90    FOR J=2 TO N
100   X=X*RND
110   NEXT J
120   S=-LOG(X)              ! EQ 41
130   A(V)=A(V)+EXP(-V*S/N)  ! EQ 42
140   NEXT V
150   NEXT I
160   R=1/T
170   FOR V=0 TO 20
180   A(V)=LGT(A(V)*R)
190   NEXT V
200   PLOTTER IS "GRAPHICS"
210   GRAPHICS
220   SCALE 0,20,-7,0
230   GRID 2,1
240   PENUP
250   LINE TYPE 9
260   FOR V=0 TO 20
270   PLOT V,A(V)       ! SIMULATION
280   NEXT V
290   PENUP
300   LINE TYPE 1
310   FOR V=0 TO 20
320   PLOT V,-N*LGT(1+V/N)   ! EXACT
330   NEXT V
340   PENUP
350   END
```

References

1. J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, Methuen and Co., London, 1964.

2. V. G. Hansen, "Importance Sampling in Computer Simulation of Signal Processors," Comput. and Elect. Engng., vol. 1, no. 4, Pergamon Press, pages 545-550, 1974.

3. Handbook of Mathematical Functions, U. S. Dept. of Commerce, National Bureau of Standards, Applied Math. Series No. 55, U. S. Gov't. Printing Office, June 1964.

4. A. H. Nuttall, "Numerical Evaluation of Cumulative Distribution Functions Directly from Characteristic Functions," NUSL Report No. 1032, 11 August 1969.

5. A. H. Nuttall, "Alternate Forms and Computational Considerations for Numerical Evaluation of Cumulative Probability Distributions Directly from Characteristic Functions," NUSC Report No. NL-3012, 12 August 1970.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
ASN (RE&S) (D. E. Mann)	1
OUSDR&E (Research & Advanced Technology (W. J. Perry)	2
Deputy USDR&E (Res & Adv Tech) (R. M. Davis)	1
OASN (Dr. R. Hoglund)	1
ONR, ONR-100, -200, -102, -222, 486	5
CNO, OP-098, -96	2
CNM, MAT-08T, -08T2, SP-20	3
DIA, DT-2C	1
NAV SURFACE WEAPONS CENTER, WHITE OAK LABORATORY	1
NRL	1
NRL, USRD	1
NORDA (Dr. R. Goodman,110)	1
USOC, Code 241, 240	2
SUBASE LANT	1
NAVSUBSUPACNLON	1
OCEANAV	1
NAVOCEANO, Code 02	1
NAVELECSYSCOM, ELEX 03	1
NAVSEASYSCOM, SEA-003	1
NAVAL SEA SYSTEM DETACHMENT, Norfolk	1
NASC, AIR-610	1
NAVAIRDEVCEN	1
NOSC	1
NOSC, Code 6565 (Library)	1
NAVWPNSCEN	1
DTNSRDC	1
NAVCOASTSYSLAB	1
CIVENGRLAB	1
NAVSURFWPNCEN	1
NUWES. San Diego	1
NUWES, Hawaii	1
NISC	1
NAVSUBSCOL	1
NAVPGSCOL	1
NAVWARCOL	1
NETC	1
NAVTRAEQUIPCENT (Technical Library)	1
APL/UW, Seattle	1
ARL/PENN STATE, State College	1
CENTER FOR NAVAL ANALYSES (ACQUISITION UNIT)	1
DTIC	12
DARPA	1
NOAA/ERL	1
NATIONAL RESEARCH COUNCIL	1
WEAPON SYSTEM EVALUATION GROUP	1
WOODS HOLE OCEANOGRAPHIC INSTITUTION	1
ARL, UNIV OF TEXAS	1
MARINE PHYSICAL LAB, SCRIPPS	1
Dr. David Middleton, 127 East 91st St., New York, NY 10028	1

INITIAL DISTRIBUTION LIST (Cont'd)

Addressee	No. of Copies
Herbert Gish, Bolt, Beranek, and Newman, 50 Moulton St., Cambridge, MA 02138	1
Prof. Louis Scharf, Dept. of Electrical Engin. Colorado State University, Fort Collins, CO 80523	1
Prof. Donald Tufts, Dept. of Electrical Engin. University of Rhode Island, Kingston, RI 02881	1
C. Hindman, TRW, Defense & Space Systems Group, One Space Park, Redondo Beach, CA 90278	1
Dr. S. L. Marple, Wash. Systems Engin. Div., The Analytic Sciences Corp, McLean, VA 22102	1
T. E. Barnard, Chesapeake Instrument Div., Gould Inc., Glen Burnie, MD 21061	1
Prof. P. M. Schultheiss, Dept. of Electrical Engin. P.O. Box 2157, Yale University, 15 Prospect Street, New Haven, CT 06520	1
Dr. Allan G. Piersol, Bolt, Beranek, and Newman, 21120 Van Owen Street, P.O. Box 633, Canaga Park, CA 92305	1
Prof. Y. T. Chan, Dept. of Electrical Engin., Royal Military College, Kingston, Ontario, Canada k7L 2W3	1
Frank W. Symons, Applied Research Laboratory, Penn State University, P.O Box 30, State College, PA 16801	1
Donald G. Childers, Dept. of Electrical Engin., University of Florida, Gainesville, Florida 32611	1

F

7